

- Durée : 2h
- Documents autorisés : cours et TD machine
- **A rendre : pour chaque exercice, une archive contenant le script Matlab avec toutes les fonctions nécessaires pour l'exécuter**

1 Dataviz

(5 points)

La matrice $X \in \mathbb{R}^{435 \times 16}$ du fichier `elections.mat` contient des informations sur le vote de 435 députés américains sur 16 sujets différents (handicaps, éducation, sécurité sociale, aides internationales ...). Ces votes peuvent être 1 (pour), -1 (contre), 0 (abstention). La variable Y représente le parti politique (-1 : démocrates, 1 : républicains) du votant.

L'objectif ici est de visualiser ces données en deux dimensions en utilisant une ACP.

1. Ecrire un programme matlab qui réalise les fonctionnalités suivantes :
 - calculer la matrice de projection P de l'ACP
 - Projeter les données en deux dimensions en utilisant P
 - Visualiser la projection en 2D avec une couleur/un symbole différent pour chaque type de votant (démocrate ou républicain)

2 Savoir reconnaître les stars

(15 points)

On dispose de données astronomiques sur les naines blanches (label $y = -1$) et les quasars (label $y = 1$). Les données consistent en 4 variables $\mathbf{x}_{(1)}, \dots, \mathbf{x}_{(4)}$ représentant la différence de luminosité des objets dans différentes longueurs d'onde. Les fichiers `astroapp.mat`, `astroval.mat`, `astrotest.mat` contiennent ces variables ainsi que les labels des points et serviront respectivement pour l'apprentissage, la validation et le test.

On cherche un modèle SVM de la forme

$$f(x) = \mathbf{w}^\top \phi(\mathbf{x}) + b$$

où le vecteur $\phi(\mathbf{x})$ est formé par les variables initiales $\mathbf{x}_{(j)}$ et leurs carrés $\mathbf{x}_{(j)}^2$ c'est-à-dire

$$\phi(\mathbf{x})^\top = \left(\mathbf{x}_{(1)} \quad \mathbf{x}_{(2)} \quad \mathbf{x}_{(3)} \quad \mathbf{x}_{(4)} \quad \mathbf{x}_{(1)}^2 \quad \mathbf{x}_{(2)}^2 \quad \mathbf{x}_{(3)}^2 \quad \mathbf{x}_{(4)}^2 \right)$$

1. Ecrire un programme Matlab qui réalise les opérations suivantes :
 - A partir des données d'apprentissage $X_a \in \mathbb{R}^{N_a \times 4}$, former la matrice $\Phi_a^{N_a \times 8}$ dont les colonnes sont les variables $\mathbf{x}_{(j)}$ et leurs carrés $\mathbf{x}_{(j)}^2$. Faire de même pour les ensembles de validation et de test.
 - apprendre un modèle SVM linéaire $f(\mathbf{x}) = \mathbf{w}^\top \phi(\mathbf{x}) + b$. On mettra en évidence la sélection du "meilleur modèle".
2. Évaluer le modèle sur les données d'apprentissage et de test en affichant les matrices de confusion et les erreurs de classification respectives.